PAPER • **OPEN ACCESS**

Mining association rule based on the diseases population for recommendation of medicine need

To cite this article: M Harahap *et al* 2018 *J. Phys.: Conf. Ser.* **1007** 012017

Mining association rule based on the diseases population for recommendation of medicine need

M Harahap*, A M Husein, S Aisyah, F R Lubis, B A Wijaya

Faculty of Technology and Computer Science, Universitas Prima Indonesia, Indonesia

E-mails: *mawaddah@unprimdn.ac.id, amirmahmud@unprimdn.ac.id

Abstract. Selection of medicines that is inappropriate will lead to an empty result at medicines, this has an impact on medical services and economic value in hospital. The importance of an appropriate medicine selection process requires an automated way to select need based on the development of the patient's illness. In this study, we analyzed patient prescriptions to identify the relationship between the disease and the medicine used by the physician in treating the patient's illness. The analytical framework includes: (1) patient prescription data collection, (2) applying k-means clustering to classify the top 10 diseases, (3) applying Apriori algorithm to find association rules based on support, confidence and lift value. The results of the tests of patient prescription datasets in 2015-2016, the application of the k-means algorithm for the clustering of 10 dominant diseases significantly affects the value of trust and support of all association rules on the Apriori algorithm making it more consistent with finding association rules of disease and related medicine. The value of support, confidence and the lift value of disease and related medicine can be used as recommendations for appropriate medicine selection. Based on the conditions of disease progressions of the hospital, there is so more optimal medicine procurement.

1. Introduction

The hospital is one of the health healing and recovery facilities for the patient [1]. Most hospitals in the implementation of health care activities have implemented Hospital Management Information System (HMIS), so in health services to patients have been recorded in the database, ranging from registration of the process of payment of health costs. However, the implementation of the process of medicine management in the hospital pharmacy installation is still not optimal because there is still often a vacancy at medicine stock. Medication is a good need for sick people with 50-60% of the overall budget of the hospital which is used at medication and medical equipment. Management of non-optimal medicine handling will adversely affect the hospital both medically and economically [2].

Selection of medicine need is the first phase at the cycle planning of procurement medicine [3] in [2]. The process of medicine selection is based on the patient's disease population. The patient's disease population may change which based on the patient's diagnostic volumes of data stored in the database, so it is need to require an automated way to select medicine requirement based on disease progression. Data mining is one of useful technique for extract and defining patterns of datasets in databases into information [4]. Application of data mining methods of the health field is proposed by many researchers, such as predicting heart disease [5], [6], health insurance [7], disease identification

[6], hypertension [8], insurance cheating, low cost patient medical solutions, related disease detection, treatment and other treatment methods [9].

Apriori is one of the Mining Association methods of Data Mining to find all related adjustment items in a database transaction that fill minimum of rules and limits or another limit [10]. Apriori algorithm that is proposed by many researchers like finding the associaton rules of Chinese traditional medicine [11], Chinese herbal medicine [12], diagnosing patients diseases with hypertensive symptoms [8], disease identification [6], detection of heart disease factors for men and women [5], identification of symptoms with traditional Korean therapy [13]. Apriori algorithm can reduce the number of candidates that must be calculated by butchering method, it has good performance [14], with the required scanning process of each iteration will increase the high time computation.

Some researchers use comparative methods to optimize time computation of apriori algorithms, one of them is the application of the k-means clustering method and it proves to be very accurate [15], [16], [17]. K-Means Clustering is one of the hard partition classification techniques, it is efficient in grouping large and fast datasets in [18] calculations, but limited at numerical data [19].

At this research, we analyze the patient prescriptions based on a doctor's diagnosis in the hospital database to identify the relationship between the disease and the medicine used by the physician in treating the patient's illness. The application of the k-means clustering method was used to find the 10 more dominant diseases in health dataset in year 2015 and 2016, then we use an apriori algorithm to find useful relationships and information between disease and related medicine based on support, confidence and lift values. Our paper is structured as follows: section 2 of the related research, the proposed method is described in Section 3. Section 4 is results and conclusion is in Section 5.

2. Related research

Research [15], applied a combination of k-means clustering algorithms with a priori in consumer data to find association rules, from the results of the study of consumer data clustering with k-means algorithm showed a significantly better and consistent influence on a priori based on value support. So it provides useful information service providers to offer the right products / ads to the right consumer. [16] Applied on clustering algorithm is to improve the performance of a priori algorithms to find solutions by generating different items each site on a cloud-based network. [17], proposed by a novel method of clustering using k-means and a priori with the aim of allocating unique id for the object of the cluster. Each objects of the group has a certain position which may vary depending on the circumstances, the id is allocated then applied to k-means clustering method along with the a priori algorithm. [20] Applying k-means clustering is to analyze goals treatment in breast cancer based on user behavior, datasets using UCI with 569 data and 32 attributes. [11] proposed modification of the Association Rules Mining method to study the structural character of the Traditional Chinese medicine (TCM) pairs with a dataset source of 625 medicine data onto 347 medicines and 5 types of cold, hot, warm, cool and normal properties. The application of a priori algorithms is used to find out some specific medicine or properties more commonly used in medicine pairs by comparing a priori by proposed method, from the test results based on statistical tests, optimal proposed method finding association rules on medicine than previous methods.

[21] proposed the mining of health data to find the pattern of illness that occurs to patients by seeking symptom relation of disorders in the medical database. Yan Yan's research, Wang Chunyan, Li Min developed a multi-model based on a priori algorithm at the hospital to extract data from the database to produce useful information in medical decision making, [20] proposed a priori algorithm to find the characteristics of headache on traditional medicine, making it easier about doctors' decision on recipes for various kinds for headache sufferers.

3. Methodology

Apriori Algorithm is the most famous algorithm for finding patterns of a database that has a frequency or support above a certain threshold called the minimum support term. A priori algorithm consists of several stages of iteration, each iterations will generate a calculated frequency pattern by scanning the database to obtain support of each items, items that have support above the minimum support are selected into high frequency patterns of length one or often called 1-itemset . K-itemset is a term of a set consisting of k items. In the second iteration process will produce 2-itemset which each set has two items [22]. The rules of association are the implications of the form X → Y, where X is the antecedent and Y is the consequence of the rule. Thus X∩Y = Φ. The support of the item set is defined as the ratio of the number of transactions containing items set to the total number of transactions. Trust of association rule X → Y is the probability that Y transaction contains an association rules mining X algorithm.

Support ot the association rule X→Y :

$$Support\ (X,Y) = P(X,Y) = \frac{\{medicine\ who\ bought\ X\ and\ Y\}}{\{medicine\}} \tag{1}$$

Confidence of association rule X→Y:

$$Confidence\ (X \rightarrow Y) \equiv P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{\{medicine\ who\ bought\ X\ and\ Y\}}{\{medicine\}} \tag{2}$$

Lift, also known as interest of association rule $X \rightarrow Y$:

$$Lift\ (X \rightarrow Y) = \frac{P(X,Y)}{P(X)\ P(Y)} = \frac{P\ (X|Y)}{P\ (XY)} \tag{3}$$

K-Means algorithm is one of the most popular clustering algorithms used cause it has a simple algorithm, easy to implement and efficient in its complexity [23]. The grouping of k-means is based on proximity to each other according to the Euclidean distance. It takes k as an input parameter and partition a set of n objects (1) from k cluster. The average value of the object (2) is taken as the resemblance (3) to the parameter to form the cluster. Cluster mean or center is formed by random selection of object k. Comparing most similarities (4) of other objects is assigned to the cluster. For each data vector the algorithm calculates the distance between the data vector and each clan centroid using the equation [24]. The steps in the K-means algorithm are as follows:

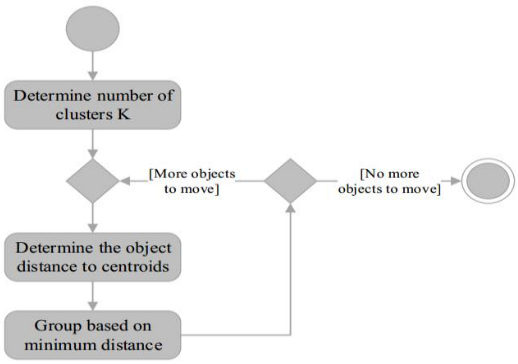

Figure 1. Data and control flow of *K*-Means algorithm.

In this study we used patient prescription datasets in 2015 and 2016 from two hospitals, we apply these datasets source to the MySQL database to facilitate the process of cleaning and transformed data. After the process of cleaning the noise data, we concluded 651.378 prescriptions with 12.015 patient data and 1.945 medicine type data for 2015, when the 2016 patient prescription dataset amounted to 956.152 prescriptions, 18.416 patients and 1.835 medicine type, like at the first table.

Table 1. Datasets source.

Year	Prescription	Patient	Medicine Type
2015	651.378	12.015	1.945
2016	956.152	18.416	1.835

The main objective of this study was to classify the 10 more dominant disease populations based on the patient's disease progression using the k-means algorithm on the patient prescription dataset. From this clustering, we apply an a priori algorithm to establish the relationship between disease and related medicine based on the value of support, confidence and lift. This knowledge can be a recommendation of appropriate medicine selection of the procurement in medicine to be more optimal to avoid the occurrence vacancy at stock of medicine in pharmacy hospital.

4. Results and Discussion

Tests were conducted to find the association for disease with medicine in patient prescription dataset, on dataset 2015 and 2016 used, we made as material Analysis and identification process of patients disease pattern. The initial step of the patient's prescription dataset will be grouped with the k-means algorithm of the dataset by 2015 consisting of 651, 378 prescriptions, 12,015 patients and 1,945 medicines, when the 2016 patient prescription dataset consists of 965,152 prescriptions, 18,416 patients and 1,835 medicaments. In grouping the disease, we use 3 (three) variables, those are age, gender and disease. Gender (man and women), age (infant, toddler, children, adult and elderly) the disease variables use ICD10 with 21,591 kinds of diseases. In table 1 is the result of cluster and number of instances, table 2 results cluster of attributing to patient prescription dataset 2015 consisting of 10 clusters.

Table 2. Cluster dataset 2015.

	# of instance	ratio
Cluster0	816	15%
Cluster1	753	14%
Cluster2	490	9%
Cluster3	487	9%
Cluster4	437	8%
Cluster5	369	7%
Cluster6	351	6%
Cluster7	306	6%
Cluster8	1,193	22%
Cluster9	207	4%

Table 3. Result attribute dataset 2015.

Age	Gender	Disease Code
Elderly	Women	L10
Adult	Women	A15
Adult	Women	K30
Adult	Women	E11
Adult	Women	O000
Elderly	Women	M15
Elderly	Man	I69.3
Elderly	Man	A01
Elderly	Women	H26
Elderly	Women	M53

In table 1 is the result of grouping of dataset year 2015 with three attributes that is age, gender and disease code ICD10. The result of grouping of three attributes is seen in table 1. In cluster has the highest value that is 22% with Elderly age attribute, gender women and disease code H26.9 compared with 10 other diseases. The results of the clustering of 10 diseases in 2016 dataset are six similar diseases of the 2015 dataset and four new diseases, but disease code H26.9 has the highest 40% compared with 9 other diseases. This provides information that cataract disease is the highest disease suffered by patients in 2015 and 2016. Results of grouping dataset in 2016 are as in table 3 and table 4.

Table 4. Cluster dataset 2016.

	# of instance	Ratio
Cluster0	1,250	40%
Cluster1	377	12%
Cluster2	298	10%
Cluster3	126	4%
Cluster4	92	3%
Cluster5	85	3%
Cluster6	387	12%
Cluster7	222	7%
Cluster8	208	7%
Cluster9	78	2%

Table 5. Result cluster attribute dataset 2016.

Age	Gender	Disease Code
Elderly	Women	H269
Adult	Women	A151
Adult	Women	H612
Adult	Man	H661
Elderly	Man	A012
Adult	Women	H608
Adult	Women	EL1
Elderly	Women	I10
Adult	Women	J30.3
Adult	Women	O000

Based on the results of the grouping of 10 of the highest diseases in the dataset of 2016, we used association rules to find the relationship between medicine-related illnesses by forming binary matrices in which columns were medicine and rows represented 10 of the highest diseases and each cells had 0 and 1. We analyzed without considering the dosage and the way of the medicine used due to varying doses. Table 5 shows the support, trust, adoption of association rules between the top ten diseases (antecedent) and the related medicine(consequent) which has a minimum limit value of Support 20% and confidence 65%. Trust and lift values can be used to assess the rules of association. Medicine that has high confidence and lifting values has a relationship of diseases such as illness with ICD code H26.9 have a trust value and support for the medicine Ciprofloxacin 500 Mg Tablet means Ciprofloxacin 500 Mg medicine the tablet is most commonly used for unspecified Cataract disease, but the value of lift to this medicine is relatively low. This implies that the Ciprofloxacin 500 Mg Tablet medicine is often used in other diseases, the Polidemisin Eye Drop medicine has a high lift value, it means that the Polidemisin Eye Drop medicine is a special remedy for unspecified Cataract disease.

Table 6. Association rules between diseases and related medicine (Min Sup: 20%, Min Conf : 65%).

Antecedent (disease Code)	Disease name	Consequent (Medicine)	Support	Confidence	Lift
H26.9	UNSPECIFIED CATARACT	CIPROFLOXACIN 500 MG TABLET	3.5	48.2	1.6
		CENDO VITROLENTA MINIDOSE	4.7	31.9	3.7
		CENDO CATARLENT MINIDOSE	2.6	76.5	2.6
		POLIDEMISIN EYE DROP	2.5	70.4	4.2
		VIGAMOX TETES MATA	4.1	48.2	2.8
		NEVANAC 0,1%	2.7	86.2	2.6
		CIPROFLOXACIN 500 MG TABLET	2.2	71.3	2.5
		CRAVIT 5ML	2.7	67.6	3.4
		FLAMAR TETES MATA	2.3	80.5	3.2
A15.1	TUBERCULOSIS OF LUNG	ALPRAZOLAM 0.5 MG TABLET	2.6	59.5	2.8
		BECOM C KAPLET	2.4	85.9	2.3
		NATRIUM DIKLOFENAC 50 MG TABLET	2.4	63.8	1.1
		LEVOFLOXACIN FC 500 MG TABLET	2.6	47.3	2.2
		MELOXICAM 15 MG TABLET	3.2	27.2	2.2

NOVALGIN INJEKSI	2.9	75.6	2.1
OMEPRAZOLE TABLET	2.9	24.4	1.5
OPICEF 500 MG KAPSUL	2.8	41.9	2.4
PARACETAMOL TABLET	2.1	57.8	2.2
RANITIDINE 25 MG INJEKSI	2.9	57.1	2.2
TEGADERM PAD 9X25 CM	2.3	44.4	1.5
TRAMIFEN TABLET	2.7	68.9	2.1
GENTAMYCIN 0.1% SALEP	2.7	69.3	2.3
METRONIDAZOLE 500 MG TABLET	2.3	89.9	0.3
CEFADROXIL 500 MG KAPSUL	2.6	52.1	2.1
SARUNG TANGAN NON STERIL (M)	2.2	47.1	2.1
MASKER EARLOOP (KARET)	2.8	48.2	1.1
SPUIT 3 CC	3.1	54.3	1.3
SPUIT 5 CC	2.5	34.1	1.3
SPUIT 10 CC	2.2	54.1	2.7
CEFTRIAXONE 1 GR INJEKSI	2.5	36.5	1.2
ALCOHOL SWAB	3.2	55.3	1.4
ANSEL GAMEX NO 7,5	2.4	85.1	1.4
CIPROFLOXACIN 500 MG TABLET	2.5	55.4	1.1
HI-BONE CAPSUL	3.4	40.3	2.9
VERBAND GULUNG SWALOW 15 CM	2.6	42.2	2.7
ASAM MEFENAMAT 500 MG KAPLET	2.5	60.2	1.7
LOMATUELL	2.9	33.8	1.3
TEGADERM PAD 9X10 CM	2.6	59.7	1.2
SARUNG TANGAN NON STERIL (S)	2.9	38.8	2.2
CEBEX KAPSUL	3.7	52.7	2.2
NEURODEX TABLET	0.6	37.4	2.1
GENTAMYCIN INJEKSI	1.4	84.3	2.4
VITAMIN B KOMPLEX	2.2	75.9	1.5
CO AMOXICLAV TABLET	1.1	36.6	1.6
VITAMIN C TABLET	3.9	88.3	2.2
CEFADROXIL 125MG/60 ML SIRUP	1.8	71.1	1.4
CALCIUM LACTAT TABLET	0.6	63.2	2.2
METHYLPREDNISOLONE 4 MG TABLET	1.9	50.4	1.5
MELOXICAM 7,5 MG TABLET	1.3	61.6	2.1
LEVOFLOXACIN FC 500 MG TABLET	2.4	25.7	2.6
CEFIXIME 100 MG KAPSUL	3.1	31.9	2
KETOROLAC 3% INJEKSI	2.4	24.4	2.3
ONDANSETRON 4 MG INJEKSI	2.7	67.3	2.2
ASAM MEFENAMAT 500 MG KAPLET	1.1	22.9	2.9
RANITIDINE 150 MG TABLET	3.2	89.5	2.1
VITAMIN C TABLET	1.9	78.9	1.3
ETABION (NOVABION) TABLET	3.1	80.6	1.7
CTM TABLET	1.6	42.6	1.1
AMOXICILLIN 500 MG TABLET	0.2	34.7	1.3
CIPROFLOXACIN 500 MG TABLET	2.9	42.5	1.9
METRONIDAZOLE 500 MG TABLET	2.2	34.4	2.2
CEFADROXIL 500 MG KAPSUL	1.1	67.8	2.1
CALCIFAR PLUS(KALSIUM LAKTAT) KAPLET	1.6	60.9	2.1
GENTAMYCIN 0.1% SALEP	1.5	38.6	1.1
PARACETAMOL TABLET(PACETIK TABLET)	1.5	81.8	2.8
MECOBALAMIN 500 MG KAPSUL	1.3	60.4	1.4
STERIL WATER 25 ML OTSU	3.1	34.6	1.4
NACL 500 ML WIDA	3.2	22.4	2.7
RINGER LACTAT 500 ML WIDA	3.8	31.4	1.2
RANITIDINE 25 MG INJEKSI	0.4	78.4	1.7

		SPUIT 3 CC	2.7	35.9	1.5
H61.2	IMPACTED CERUMEN	CETIRIZINE 10 MG TABLET	2.1	28.9	0.9
		METHYLPREDNISOLONE 4 MG TABLET	1.9	57.3	0.9
		H2O2 50%	1.9	27.9	0.6
		BOTOL 30 CC	0.5	64.9	1.9
		LEVOFLOXACIN FC 500 MG TABLET	2.2	85.2	1.6
		PIPET TETES KACA	1.5	33.5	1.1
		CEFADROXIL 500 MG KAPSUL	1.2	30.2	1.4
		CIPROFLOXACIN 500 MG TABLET	1.4	71.8	0.7
		LANSOPRAZOLE 30 MG TABLET	2.8	55.1	0.6
		METRONIDAZOLE 500 MG TABLET	0.4	82.4	1.3
		OMEPRAZOLE TABLET	1.2	24.1	0.7
		NATRIUM DIKLOFENAC 50 MG TABLET	1.9	80.6	0.8
		AVAMYS NASAL SPRAY	1.6	45.1	1.6
		CTM TABLET	1.8	24.3	1.5
		VITAMIN B KOMPLEX	2.7	89.5	0.7
		ILIADIN SPRAY 0.05% DEWASA	2.3	68.9	1.2
		SARUNG TANGAN NON STERIL (M)	2.4	52.8	1.5
		PARACETAMOL TABLET(PACETIK TABLET)	2.7	70.9	1.5
		CEFIXIME 200 MG KAPSUL	3.3	41.1	0.6
		MASKER EARLOOP (KARET)	0.4	67.3	1.1
		MECOBALAMIN 500 MG KAPSUL	2.4	81.3	0.2
		ILIADIN SPRAY 0.05% DEWASA	0.5	64.2	0.5
		AZITROMYCIN 500 MG TABLET	0.1	88.7	0.5
		CEFIXIME DRY SYRUP	0.7	24.7	0.3
H66.1	CHRONIC TUBOTYMPANIC SUPPURATIVE OTITIS MEDIA	CETIRIZINE 10 MG TABLET	4.4	26.2	3.8
		LEVOFLOXACIN FC 500 MG TABLET	3.9	73.4	3.4
		METHYLPREDNISOLONE 4 MG TABLET	4.8	85.4	1.9
		METRONIDAZOLE 500 MG TABLET	4.9	59.6	0.1
		CIPROFLOXACIN 500 MG TABLET	4.5	71.3	3.2
		H2O2 50%	4.8	24.9	2.8
		CTM TABLET	3.4	73.6	1.9
		BOTOL 30 CC	4.4	72.8	0.5
		PIPET TETES KACA	4.5	37.4	3.8
		DEXAMETHASON TABLET (DIOMETA-FLACOID)	4.3	34.5	0.5
		CLINDAMYCIN 300 MG KAPSUL	.3.0	59.7	1.6
A01.2	PARATYPHOID FEVER B	CONCOR 2.5 MG TABLET	2.1	61.5	1.7
		FUROSEMIDE 40 MG TABLET	0.8	67.3	1.9
		APTOR TABLET	2.0	27.2	1.3
		SPIRONOLACTONE 25 MG TABLET	1.7	31.2	0.7
		MICARDIS 80 MG TABLET	2.2	40.3	3.9
		NITROCAF RETARD FORTE	2.2	48.4	2.2
		CLOPIDOGREL 75 MG TABLET	0.8	82.3	0.2
		MICARDIS 40 MG TABLET	2.4	66.2	1.2
		SIMVASTATIN 20 MG TABLET	0.7	27.1	3.5
		ISOSORBIDE DINITRAT 5 MG TABLET	1.6	57.1	1.0
		ADALAT OROS 30 MG TABLET	1.6	72.9	3.5
		ALPRAZOLAM 0.5 MG TABLET	2.0	42.7	1.6
		RAMIPRIL 5 MG TABLET	0.4	21.3	0.4
		APTOR TABLET	1.9	59.3	2.5
H60.8	OTHER OTITIS EXTERNA	CETIRIZINE 10 MG TABLET	1.3	87.6	3.6
		LEVOFLOXACIN FC 500 MG TABLET	1.7	89.2	3.8
		METHYLPREDNISOLONE 4 MG TABLET	1.5	29.9	3.5
		H2O2 50%	0.7	47.5	3.6
		BOTOL 30 CC	0.1	78.3	3.3
		SARUNG TANGAN NON STERIL (M)	0.2	86.7	3.4

		PIPET TETES KACA	0.9	35.2	3.2
		MASKER EARLOOP (KARET)	1.6	49.4	3.6
		OMEPRAZOLE TABLET	1.6	75.9	3.8
E11	TYPE 2 DIABETES MELLITUS	METFORMIN 500 MG TABLET	1.8	88.9	0.8
		ALCOHOL SWAB	2.0	30.1	0.7
		GLIMEPIRIDE 2 MG TABLET	1.9	25.3	1.7
		DIAMICRON MR 60 MG TABLET	0.1	49.3	1.0
		MICARDIS 80 MG TABLET	1.9	78.4	0.7
		OMEPRAZOLE TABLET	0.3	43.6	1.1
		GABAPENTIN 300 MG KAPSUL	1.1	73.9	1.4
		GLIQUIDONE TABLET	2.0	81.7	0.2
		ANTASIDA DOEN 60 ML SYRUP	1.7	41.4	1.4
		AMLODIPINE BESYLATE 10 MG TABLET	0.7	36.9	1.1
		MECOBALAMIN 500 MG KAPSUL	1.8	63.6	1.7
		APTOR TABLET	1.5	37.9	0.1
		ACARBOSE 50 MG TABLET	1.0	54.8	1.5
		MELOXICAM 15 MG TABLET	0.2	21.1	0.5
		GLUCOSAMINE TABLET 500 MG	0.2	42.5	1.6
		ALPRAZOLAM 0.5 MG TABLET	1.9	58.7	0.7
		SIMVASTATIN 20 MG TABLET	1.6	80.5	0.7
		SIMVASTATIN 10 MG KAPLET	1.6	80.8	0.7
		MICARDIS 40 MG TABLET	0.3	47.4	0.6
		LANSOPRAZOLE 30 MG TABLET	2.1	54.1	0.3
		AMITRIPTYLINE TABLET	0.6	75.9	0.8
		CANDESARTAN 8 MG TABLET	0.6	64.9	1.0
		NOVORAPID	1.5	68.1	1.1
		ADALAT OROS 30 MG TABLET	1.5	43.1	1.7
		CURCUMA FCT TABLET	1.0	59.9	0.8
		CILOSTAZOL 100MG TABLET	0.1	58.7	2.0
		CLOPIDOGREL 75 MG TABLET	0.1	71.8	1.8
		BETAHISTIN MESILAT 6 MG TABLET	1.4	55.2	0.1
		CETIRIZINE 10 MG TABLET	0.4	65.1	1.4
		GABAPENTIN 300 MG KAPSUL	1.7	69.1	1.3
		ETABION (NOVABION) TABLET	0.4	43.2	1.7
		SIMVASTATIN 10 MG KAPLET	1.4	53.3	0.6
		GLIMEPIRIDE 2 MG TABLET	1.3	50.5	1.9
		RETAPHYL SR 300 MG TABLET	0.8	32.1	1.2
		FLUNARIZINE TAB(SEREMIG)	1.4	76.1	2.1
		DIAMICRON MR 60 MG TABLET	1.2	30.8	0.8
		EPERISONE HCL 50 MG TABLET	1.7	61.6	0.2
		CONCOR 2.5 MG TABLET	1.7	30.7	0.5
		VALESCO 80 MG KAPSUL	1.3	79.3	1.5
		AMLODIPINE 5 MG TABLET	0.1	48.7	0.6
		LEVOFLOXACIN FC 500 MG TABLET	0.4	73.2	1.4
		CANDESARTAN 16 MG TABLET	0.2	34.2	0.8
		NOVOMIX	0.9	65.2	2.3
		KETOPROFEN 100 MG TABLET	1.1	66.7	1.1
		CAMIDRYL EXP 60 ML SYRUP	0.7	59.3	2.4
I10	ESSENTIAL (PRIMARY) HYPERTENSION	MICARDIS 80 MG TABLET	0.2	78.2	4.7
		AMLODIPINE BESYLATE 10 MG TABLET	0.8	65.3	5.4
		OMEPRAZOLE TABLET	0.9	67.1	4.8
		ADALAT OROS 30 MG TABLET	0.2	48.5	4.9
		SIMVASTATIN 20 MG TABLET	0.7	81.5	5.8
		CONCOR 2.5 MG TABLET	1.5	58.7	4.5
		ANTASIDA DOEN 60 ML SYRUP	0.3	67.2	5.1
		GLUCOSAMINE TABLET 500 MG	1.2	24.9	1.3

		MELOXICAM 15 MG TABLET	1.2	30.8	4.3
		MICARDIS 40 MG TABLET	3.9	57.1	4.6
		ALPRAZOLAM 0.5 MG TABLET	1.2	35.1	4.5
		LANSOPRAZOLE 30 MG TABLET	1.9	42.5	4.8
		APTOR TABLET	2.1	22.7	5.5
		CANDESARTAN 8 MG TABLET	3.0	49.9	4.6
		CANDESARTAN 16 MG TABLET	3.8	86.2	5.2
		SIMVASTATIN 10 MG KAPLET	3.1	54.7	4.8
		EPERISONE HCL 50 MG TABLET	3.3	38.8	4.4
		GABAPENTIN 300 MG KAPSUL	3.2	36.9	4.1
		AMLODIPINE 5 MG TABLET	2.3	31.4	4.1
		SIMVASTATIN 10 MG KAPLET	3.5	21.1	5.3
		VALESCO 80 MG KAPSUL	1.6	51.1	4.2
		MECOBALAMIN 500 MG KAPSUL	2.8	50.9	4.9
		BETAHISTIN MESILAT 6 MG TABLET	1.4	84.5	5.9
		ULSAFATE SUSPENSI 100 ML	1.9	36.2	4.2
		BISOPROLOL FUMARATE 5 MG TABLET	2.5	45.4	4.9
		METHYLPREDNISOLONE 4 MG TABLET	3.1	44.4	4.6
		ALCOHOL SWAB	3.1	44.7	5.3
		VALESCO 160 MG KAPSUL	2.2	36.9	4.3
		FLUNARIZINE TAB(SEREMIG)	3.8	60.5	5.1
J30.3	OTHER ALLERGIC RHINITIS	CETIRIZINE 10 MG TABLET	1.8	73.1	1.4
		METHYLPREDNISOLONE 4 MG TABLET	3.1	22.2	2.1
		AVAMYS NASAL SPRAY	3.9	77.4	1.2
		CTM TABLET	3.2	55.8	2.1
		SARUNG TANGAN NON STERIL (M)	2.3	29.7	1.3
		OMEPRAZOLE TABLET	3.3	60.1	1.6
		ILIADIN SPRAY 0.05% DEWASA	2.2	46.1	1.1
		MASKER EARLOOP (KARET)	2.5	23.6	1.6
		CEFADROXIL 500 MG KAPSUL	4.7	50.7	1.9
		LANSOPRAZOLE 30 MG TABLET	3.5	82.1	1.1
		ILIADIN SPRAY 0.05% DEWASA	4.3	63.3	2.4
		PARACETAMOL TABLET(PACETIK TABLET)	3.9	27.7	2.7
		TREMENZA TABLET	3.6	83.5	1.1
		NASACORT AQ	2.2	89.3	1.8
O00.0	ABDOMINAL PREGNANCY	PROMAVIT KAPSUL	2.3	79.8	2.9
		ETABION (NOVABION) TABLET	4.5	37.4	3.9
		ASAM MEFENAMAT 500 MG KAPLET	4.4	31.8	3.8
		ABBOCATH NO 18 TERUMO	4.5	51.3	2.6
		INFUSION SET ADULT (EASY FUSION VENTED)	4.6	82.2	3.8
		3-WAY BD CONNECTA PLUS 3 WHITE	4.2	67.5	2.8
		STERIL WATER 25 ML OTSU	4.9	22.6	3.3
		CEFADROXIL 500 MG KAPSUL	4.7	79.3	2.5
		CEFTRIAXONE 1 GR INJEKSI	4.6	22.9	4,4
		SPUIT 10 CC	4.4	80.4	3.5
		GILLETTE GOAL II	2.5	69.1	2.7
		LOMATUELL	3.3	78.1	3.8
		PEMBALUT WANITA BIO PANCA	3.4	81.8	2.8
		TRAMADOL 100MG/2ML INJEKSI	3.1	70.2	2.8
		SPUIT 3 CC	4.4	22.6	2.7
		NEOLUS DISP NEEDLE NO.26 TERUMO	2.4	60.1	3.4
		METRONIDAZOLE 500 MG TABLET	4.9	39.8	3.4
		SARUNG TANGAN NON STERIL (M)	4.8	38.9	3.8
		CALCIFAR PLUS(KALSIUM LAKTAT) KAPLET	4.9	24.8	3.2
		RINGER LACTAT 500 ML WIDA	3.4	63.6	2.1
		PROFENID 100 MG SUP	4.5	44.4	2.3

5. Conclusion
The application of a priori algorithms in this study aims to extract useful information about the patient prescription database sourced from two different hospitals. We use Association Rules to find the relationship between disease and related medicine based on the grouping disease using k-means algorithms. From the results of the tests, the k-means algorithm accurately classifies 10 dominant diseases in patient prescription datasets in 2015 and 2016, thus significantly affect the a priori algorithm, it is more consistently to find association rules between disease and related medicine. The value of support, confidence and lift between medicine related diseases can be useful as a recommendation of appropriate medicine selection based on the condition of disease progression of the patient, so the procurement of medicine in the hospital is more optimal.

References
[1] Dep Kes RI, 2004. *Keputusan Menteri Kesehatan Republik Indonesia Tentang Standar Pelayanan Farmasi Di Rumah Sakit dan Apotek*, Jakarta
[2] Satibi, (2014), Manajemen Obat di Rumah Sakit, Fakultas Farmasi Universitas Gadjah Mada Yogyakarta
[3] Quick, J.P., Rankin, J.R., Laing, R.O., O'Cornor, R.W. (2012). "Managing Drug Supply, the selection, procurement, distribution and use of pharmaceutical, third edition". Kumarin Press, Conecticus, USA
[4] Sanjeev Rao, Priyanka Gupta. 2012. "Implementing Improved Algorithm Over APRIORI Data Mining Association Rule Algorithm". IJCST Vol. 3, Issue 1. 489-493
[5] Jesmin Nahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, 2013, "Association rule mining to detect factors which contribute to heart disease in males and females, Expert Systems with Applications", 1086–1093
[6] M. Ilayaraja, T. Meyyappan, 2013, "Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm". Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22, 194-199
[7] R.J. Kuo∗, C.W. Shih, 2007, "Association rule mining through the ant colony system for National Health Insurance Research Database in Taiwan, Computers and Mathematics with Applications", 1303–1318
[8] A Mi Shin, RN, et al. 2017, "Diagnostic Analysis of Patients with Essential Hypertension Using Association Rule Mining", The Korean Society of Medical Informatics, 77-81. doi:10.4258/hir.2010.16.2.77
[9] H. C. Koh and G. Tan, "Data Mining Application in Healthcare", Journal of Healthcare Information Management, vol. 19, no. 2, (2005)
[10] Kotsiantis S, Kanellopoulos D (2006) "Association rules mining: A recent overview". GESTS International Transactions on Computer Science and Engineering 1: 71-82
[11] Shang E*, Duan J, Fan X, Tang Y and Ye L, 2017, "Association Rules Mining and Statistic Test Over Multiple Datasets on TCM Drug Pairs", International Journal of Biomedical Data Mining. Volume 6, Issue 1, 1000126. doi: 10.4172/2090-4924.1000126
[12] Pei-Shan Chien, Yu-Fang Tseng, Yao-Chin Hsu, Yu-Kai Lai and Shih-Feng Weng, 2013, BMC Complementary and Alternative Medicine, 2-7
[13] Yang DH, Kang JH, Park YB, Park YJ, Oh HS, et al. (2013) "Association Rule Mining and Network Analysis in Oriental Medicine". PLoS ONE 8(3)
[14] Badri Patel, Vijay K Chaudhari, Rajneesh K Karan, YK Rana. 2011. "Optimization of Association Rule Mining Apriori Algorithm Using ACO". International Journal of Soft Computing and Engineering (IJSCE) Volume-1, Issue-1. 24-26. ISSN: 2231-2307
[15] Nergis Yılmaz and Gülfem Işıklar Alptekin, 2013, The Effect of Clustering in the Apriori Data Mining Algorithm: A Case Study Proceedings of the World Congress on Engineering 2013 Vol III, ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IOP Conf. Series: Journal of Physics: Conf. Series **1007** (2018) 012017 doi:10.1088/1742-6596/1007/1/012017

[16] Nisha Rani, Yamini Chouhan, 2015, "Improving the Performance of Apriori Algorithm by Combining with Clustering Techniques", International Journal of Emerging Technologies in Engineering Research (IJETER), Volume 3, Issue 2, 13-15
[17] Syed Zishan Ali, Nikhil Tiwari and Sushmita Sen, 2016, "A Novel Method for Clustering using k-means and Apriori Algorithm", International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB16), 978-1-4673-9745-2
[18] Ghosh, S., Dubey, K. S. (2013). "Comparative Analysis of K-Means and Fuzzy C Means Algorithms". ((IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 4, 2013
[19] Bai, L., Liang, J., & Dang, C. (2011). "An initialization method to simultaneously find initial cluster centers and the number of clusters for clustering categorical data". Knowledge-Based Systems, 24(6), 785–795
[20] Miao Wang, Lei Zhang, Zihao Zhang, Chen Xu, Guobin Chen, Huiliang Shang, 2014. The "Application Characteristics of Traditional Chinese Medical Science Treatment on Headache Based on Data-Mining Apriori Algorithm". IEEE International Conference on Bioinformatics and Biomedicine. 153-157
[21] K. Zuhtuogullari. N. Allahverdi. 2011. An Improved Itemset Generation Approach for Mining Medical Databases. IEEE. 39-43. 978-1-61284-922-5
[22] Goswami D.N, Chaturvedi Anshu, Raghuvanshi C.S, 2010. "An Algorithm for Frequent Pattern Mining Based On Apriori",(IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 942-947
[23] Aggarwal, N., & Aggarwal, K. (2012). "Comparative Analysis of kmeans and Enhanced K-means clustering algorithm for data mining", International Journal of Scientific & Eninnering Research
[24] Saurabh Shah & Manmohan Singh "Comparison of A Time Efficient Modified K-Mean Algorithm with K-Mean and K-Medoids algorithm", International Conference on Communication Systems and Network Technologies, 2012

Acknowledgments
This research was supported by Universitas Prima Indonesia and one of private hospital in Medan city.